[0030] The 3D conversation pipeline can be implemented using different combinations of 2D display devices, 3D display devices, and/or intermediate devices. Examples of 2D display devices include PCs, mobile devices (e.g., phones, tablets, watches, etc.), televisions, or other traditional display screen technology. Examples of 3D display devices include mixed reality glasses, virtual reality headsets, laser plasma displays, projection caves, etc., which may include wireless or wired external compute sources. In various implementations, the display devices can include one or more displays such as at least one binocular display, at least one lightfield display, at least one holographic display, at least one wave front display, at least one 3D stereo displays, or any combination thereof. In some implementations, different types of devices (2D and/or 3D) can communicate with each other using the 3D conversation pipeline. In various implementations, 3D conversations using the 3D conversation pipeline can be between two or more devices that each act as sender devices for capturing images of users of that device and receiver devices for receiving representations of each of the other one or more participants of the conversation.

[0031] In some implementations, the 3D conversation pipeline is performed without utilizing servers for anything other than traditional communication (e.g., over the Internet), while in other implementations, a server or other cloud computing system can perform certain stages such as reconstruction, rendering, or parts of calibration. For example, the reconstruction stage for a 3D conversation involving three participants can be at a server that compiles the captured image data from the participants into a single 3D representation of each of the participants. As another example, an intermediate server might perform part of the filtering stage, whereby the viewpoints of the various participants are tracked in a 3D environment and used to determine which portions of the captured data from each other participant is needed for a recipient device to perform reconstruction and rendering for that viewpoint. Additional details for various pipeline configurations using different devices for different stages are discussed below in relation to FIG. **6**.

[0032] In various implementations, the 3D pipeline as a whole or configurations of various stages can be smoothly adjusted to upgrade or downgrade features according to a conversation context (e.g., based on processing or other computing resources available at the various devices, display capabilities of the various devices, available bandwidth, user preferences, etc.) For example, frame rates or resolution can be adjusted, 3D contouring (e.g., number of triangles used) can be adjusted, portions of images can be predicted using machine learning or based on previously captured data rather than being wholly representative of current captured data, portions of images can be replaced with avatar representations, or the conversation can be converted to a 2D or audio only version.

[0033] Embodiments of the disclosed technology may include or be implemented in conjunction with an artificial reality system. Artificial reality or extra reality (XR) is a form of reality that has been adjusted in some manner before presentation to a user, which may include, e.g., a virtual reality (VR), an augmented reality (AR), a mixed reality (MR), a hybrid reality, or some combination and/or derivatives thereof. Artificial reality content may include completely generated content or generated content combined with captured content (e.g., real-world photographs). The artificial reality content may include video, audio, haptic feedback, or some combination thereof, any of which may be presented in a single channel or in multiple channels (such as stereo video that produces a three-dimensional effect to the viewer). Additionally, in some embodiments, artificial reality may be associated with applications, products, accessories, services, or some combination thereof, that are, e.g., used to create content in an artificial reality and/or used in (e.g., perform activities in) an artificial reality. The artificial reality system that provides the artificial reality content may be implemented on various platforms, including a head-mounted display (HMD) connected to a host computer system, a standalone HMD, a mobile device or computing system, a "cave" environment or other projection system, or any other hardware platform capable of providing artificial reality content to one or more viewers.

[0034] "Virtual reality" or "VR," as used herein, refers to an immersive experience where a user's visual input is controlled by a computing system. "Augmented reality" or "AR" refers to systems where a user views images of the real world after they have passed through a computing system. For example, a tablet with a camera on the back can capture images of the real world and then display the images on the screen on the opposite side of the tablet from the camera. The tablet can process and adjust or "augment" the images as they pass through the system, such as by adding virtual objects. "Mixed reality" or "MR" refers to systems where light entering a user's eye is partially generated by a computing system and partially composes light reflected off objects in the real world. For example, a MR headset could be shaped as a pair of glasses with a pass-through display, which allows light from the real world to pass through a waveguide that simultaneously emits light from a projector in the MR headset, allowing the MR headset to present virtual objects intermixed with the real objects the user can see. "Artificial reality," "extra reality," or "XR," as used herein, refers to any of VR, AR, MR, or any combination or hybrid thereof.

[0035] While there are existing visual communication systems, they fail to enable communications comparable to in-person interaction. For example, existing video calling systems limit how much body language can be perceived, fail to provide the ability for users to move relative to each other, and introduce an intrusive layer of technology that can distract from the communication and diminish the perception of in-person communication. In addition, existing systems for providing interactions in 3D environments, such as in virtual reality chat rooms, are very computationally expensive and fail to accurately depict the communication participants. The pipeline of data capture and manipulation stages in the 3D conversation systems and processes described herein are expected to overcome these problems associated with conventional video and 3D interaction techniques and are expected to reduce computational requirements of 3D communication while providing more realistic interactions. Furthermore, by including multiple static customizations and/or dynamic, context-based selections of pipeline stages and stage configurations, the 3D conversation system can operate with devices of many more processing and display capabilities and with many more bandwidth availabilities than the existing systems. In addition, while the disclosed 3D conversation systems provide a result comparable to an in-person experience, the processes and systems to achieve this result are not analogs of existing